Health Data in Practice lecture series

Electronic health records: sources of hospital data

Carol Dezateux

13th October 2020

Institute of Population Health Sciences

wellcome

Queen Mary University of London
Barts and The London

Learning Objectives

At the end of this lecture you will be able to

- Describe the four domains of data held within Hospital Episode Statistics data for England
- Understand definitions of a spell, episode and continuous inpatient spell
- Understand the sources of equivalent data in other UK countries
- Understand the data sources and types of information available on cancer registration in England
- Understand where to obtain further information on additional sources of specialised data and patient registries
- Summarise the key limitations of using data collected for routine care for research

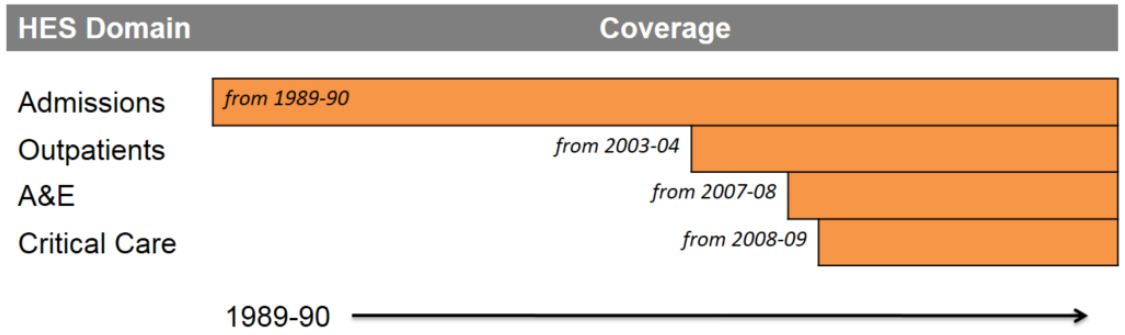
England: Hospital Episode System (HES) Data (1)

- Established in 1989 to record every 'episode' of admitted patient care delivered in England
- Primary purpose is for reimbursement of hospitals by NHS England for the care they have provided
- Generated locally, collated centrally and made available for secondary uses in annual datasets
- Covers patients cared for in NHS hospitals, private patients treated in NHS hospitals including non-English residents, admissions to private hospitals where this is funded by NHS
- Includes cross border episodes eg for Welsh residents cared for in English hospitals and vice versa
- By 2017 over 125 million records
- Additional linkages using “HESID” to eg Diagnostic Imaging Data Set etc
- The HESID is derived using a matching algorithm which looks at various combinations of the following patient identifiable fields:
  - NHS Number (fieldname NEWNHSNO)
  - Date of Birth (fieldname DOB)
  - Sex (fieldname SEX)
  - Postcode (fieldname HOMEADD)
  - Provider code (fieldname PROCODET)
  - Local patient ID (fieldname LOPATID)

Boyd A, Cornish R, Johnson L, Simmonds S, Syddall H, Westbury L, Cooper C, Macleod J. (2018). Understanding Hospital Episode Statistics (HES). Bristol, UK: University of Bristol.

England: Hospital Episode System Data (2)

- Includes four domains
  - Inpatient episodes including maternity episodes
  - Outpatient episodes
  - Accident and emergency attendance
  - Critical care

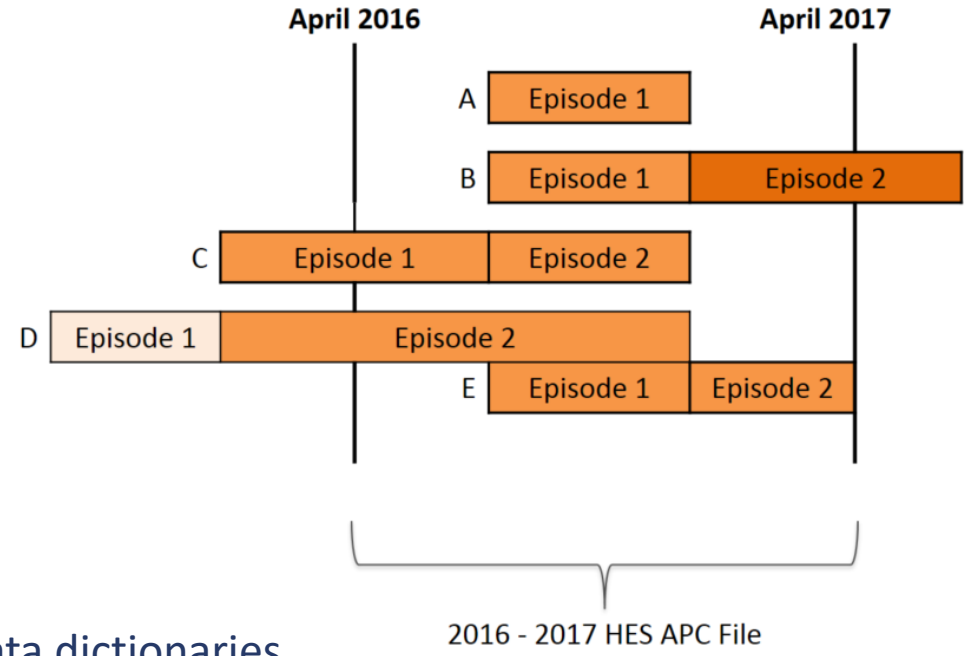


- Data are collected while patient receives treatment by each care provider and submitted on a monthly basis as a Commissioning Dataset (since 2006 using Secondary Uses Service ("SUS" data)
- SUS sends data to NHS Digital and a copy is retained in a local database which is updated incrementally
- NHS Digital carries out de-deduplication and quality assurance and deposits national dataset annually in its warehouse
- Record may be missing if record keeping error, coding error, linkage error, or prior opt out

Boyd A, Cornish R, Johnson L, Simmonds S, Syddall H, Westbury L, Cooper C, Macleod J. (2018). Understanding Hospital Episode Statistics (HES). Bristol, UK: University of Bristol.

England: Hospital Episode System Data (4)

- The HES annual dataset collates data by financial year ie 1st April to 31st March
- Episodes that straddle financial year end will be marked as unfinished in that year and finished in subsequent year



- Content documented in HES data dictionaries

https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary

Boyd A, Cornish R, Johnson L, Simmonds S, Syddall H, Westbury L, Cooper C, Macleod J. (2018). Understanding Hospital Episode Statistics (HES). Bristol, UK: University of Bristol.

England: Hospital Episode System Data (3)

Hospital Admission Records (Admitted Patient Care)

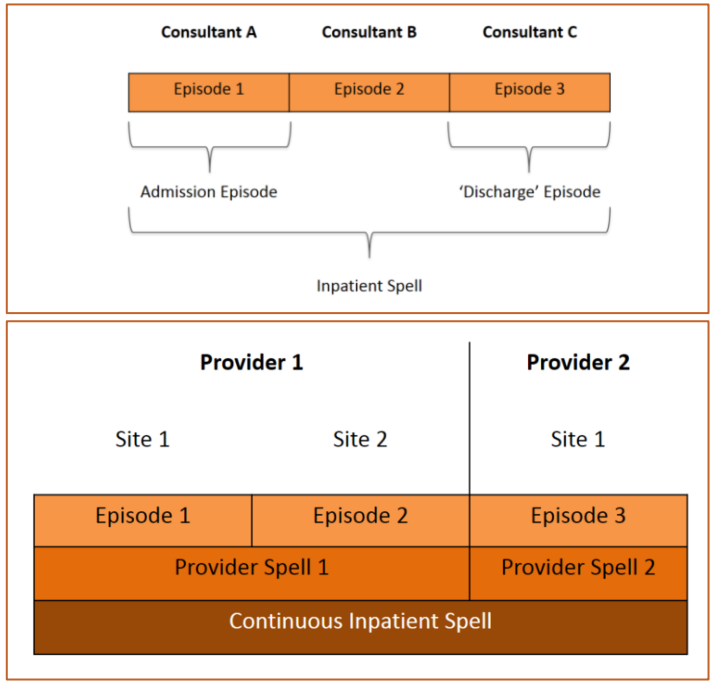
Data are recorded as **spells and episodes**

- **A spell** is a period of continuous care in one provider institution
- Within each spell there can be **episodes** defined as periods of continuous care from one nominated consultant

- Each row in HES comprises a record of a single episode
- These can be grouped into admission spells
- A spell starts when a patient is admitted
- A spell ends when they are discharged, transferred or die
- Transfer to another health care provider starts a new spell

Continuous Inpatient Spell

- all episodes and inpatient spells combined
- summarises the continuous sequence of inpatient care



Boyd A, Cornish R, Johnson L, Simmonds S, Syddall H, Westbury L, Cooper C, Macleod J. (2018). Understanding Hospital Episode Statistics (HES). Bristol, UK: University of Bristol.

England: Hospital Episode System Data (5)

Hospital Admissions records

Content documented in data dictionary
https://digital.nhs.uk/binaries/content/assets/legac_admitted_patient_care.pdf

Identifiers	Clinical Information	Demographic information	Administrative	Maternity tail
HESID (specific to each data sharing agreement)	Diagnoses and procedures (up to 20 primary and secondary)	Age (years) at admission and discharge	Method of admission (e.g. elective or emergency, birth, transfer)	Gestational Age
Episode ID	Operation Dates	Gender	Episode start and end date	Parity
Date of admission	Consultant Speciality	Index of Multiple Deprivation (IMD)	Discharge method (e.g. self-discharge, died, transferred)	Birth Weight
A&E link ID	Augmented care location	Health, Electoral and census geographies	Discharge destination (e.g. home, other destination	Maternal age
Provider details (e.g. hospital code).		Ethnic group	Time waited	Mode of delivery (e.g. forceps or spontaneous)
Registered GP practice				Birth Order (for multiple births)
				Neonatal care

Boyd A, Cornish R, Johnson L, Simmonds S, Syddall H, Westbury L, Cooper C, Macleod J. (2018). Understanding Hospital Episode Statistics (HES). Bristol, UK: University of Bristol.

Questions???

England: Hospital Episode System Data (6)

Outpatient attendances

- One row per appointment irrespective of whether attended
- Data completeness and quality high for attendance type, source of referral and main specialty and low for primary diagnosis and main procedure
- Content documented in data dictionary https://digital.nhs.uk/binaries/content/assets/legacy/pdf/4/i/hes_data_dictionary_-_outpatients.pdf

Identifiers	Clinical Information	Demographic information	Administrative
HESID (specific to each data sharing agreement)	Diagnosis (up to 12 primary and secondary diagnoses)	Age (years) at appointment	Attendance details
Appointment ID	Operative procedure(s)	Gender	Time waited
Appointment date	Consultant Speciality (e.g. Ophthalmology, Child and adolescent psychiatry)	Indices of Multiple Deprivation (IMD)	Appointment type (e.g. face-to-face, telephone)
Registered GP practice		Health, electoral and census geographies Ethnic group	

Boyd A, Cornish R, Johnson L, Simmonds S, Syddall H, Westbury L, Cooper C, Macleod J. (2018). Understanding Hospital Episode Statistics (HES). Bristol, UK: University of Bristol.

England: Hospital Episode System Data (7)

Accident and Emergency Records (A&E)

- Includes attendances (single visit) for urgent care and illness, specialty A& E depts, walk-in centres and minor injury units; first and follow up visits are distinguishable
- Documented in data dictionary https://digital.nhs.uk/binaries/content/assets/legacy/pdf/3/l/hes_data_dictionary_-_accident_and_emergency.pdf

Identifiers	Classification & clinical	Demographic information	Administrative
HESID (specific to each data sharing agreement)	Incident location (e.g. home, work, public place)	Age (years) at arrival	Arrival mode (ambulance or other)
Appointment ID	Patient group (e.g. road traffic accident, sports injury)	Gender	Attendance Category (first attendance or follow-up)
Arrival date and time	Diagnosis (up to 12 codes)	Indices of Multiple Deprivation (IMD)	Disposal (e.g. admitted, died, referred)
Registered GP practice	Anatomical area and side	Health, electoral and census geographies	Source for referral (e.g. self, GP, police, social services)
	A&E investigation (e.g. x-ray, toxicology)	Ethnic group	Visit duration

Boyd A, Cornish R, Johnson L, Simmonds S, Syddall H, Westbury L, Cooper C, Macleod J. (2018). Understanding Hospital Episode Statistics (HES). Bristol, UK: University of Bristol.

England: Hospital Episode System Data (8)

Adult Critical Care Records

- Critical care periods in designated intensive care or high dependency units requiring constant support and monitoring to maintain function in one organ
- Critical Care Minimum Dataset of up to 34 items
- Documented in data dictionary

https://digital.nhs.uk/binaries/content/assets/legacy/pdf/7/9/hes_data_dictionary_-_adult_critical_care.pdf

Identifiers	Classification & clinical	Demographic information	Administrative
HESID (specific to each data sharing agreement)	Treatment function (e.g. transplantaion surgery, burns care)	Age (years) at arrival	Admission source (e.g. same hospital, transfer)
Provider code	Critical care level and duration of care at that level	Gender	ACC Unit function (e.g. renal, neuroscience)
Start date and time	Variables indicating duration of care in specific areas (e.g. renal support)	Indices of Multiple Deprivation (IMD)	Discharge location (e.g. ward details, home)
Registered GP practice	Maximum number of organs being supported	Health, electoral and census geographies	
		Ethnic group	

R, Johnson L, Simmonds S, Syddall H, Westbury L, Cooper C, Macleod J. anding Hospital Episode Statistics (HES). Bristol, UK: University of Bristol.

Questions???

England: Hospital Episode System Data (9)

Coding systems

- Vary over time eg ICD9 ICD10
- Vary between HES datasets
- Include Procedure Codes Office of Population Censuses and Surveys (OPCS) clinical classification

Data quality

- Not clinically driven ie coded by coding clerks for billing
- Data are not timely or returned back meaningfully to clinician
- Lack of transparency over linkage algorithms used ınternally
- Although NHSD carries out many quality assurance procedures
  - 12.7% and 24.4% of primary and secondary diagnoses respectively may be incorrect
  - 15.9% and 15.3% of primary and secondary procedures respectively may be incorrect
- Not all diagnoses are included – only those related to reason for admission
- Linkage to research studies can be useful validation although gold standard unclear

Research Access

- Via application to DARS (Data Access Request Service)

Boyd A, Cornish R, Johnson L, Simmonds S, Syddall H, Westbury L, Cooper C, Macleod J. (2018). Understanding Hospital Episode Statistics (HES). Bristol, UK: University of Bristol.

Hospital episode data in other UK countries

Wales

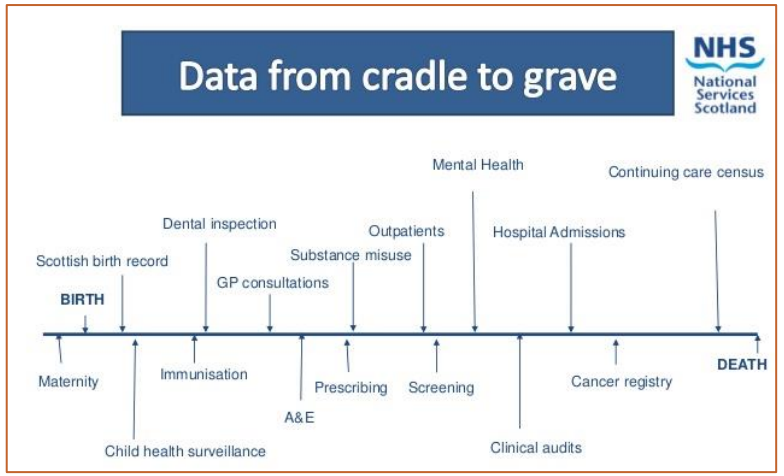
- The Patient Episode Database for Wales (PEDW) from 1999 to present
- Accessed from NHS Wales Informatics Service (NWIS) via SAIL Databank
- Full list of all datasets https://data.ukserp.ac.uk/Organisation/Category?nodeId=1&orgId=0
- Also family justice and administrative data eg census, education

Scotland

- The Scottish Morbidity Record (SMR) from 1981 to present
- Accessed from Information Statistics Division (ISD) via eDRIS (electronic Data Research & Innovation Service)
- National Data catalogue https://www.ndc.scot.nhs.uk/index.asp?Co=Y

Northern Ireland

- Northern Ireland Hospital Statistics Dataset from 2007 until present
- Access arrangements under development
- https://www.nisra.gov.uk/statistics/health-and-social-care/health-and-social-care-statistics



Questions???

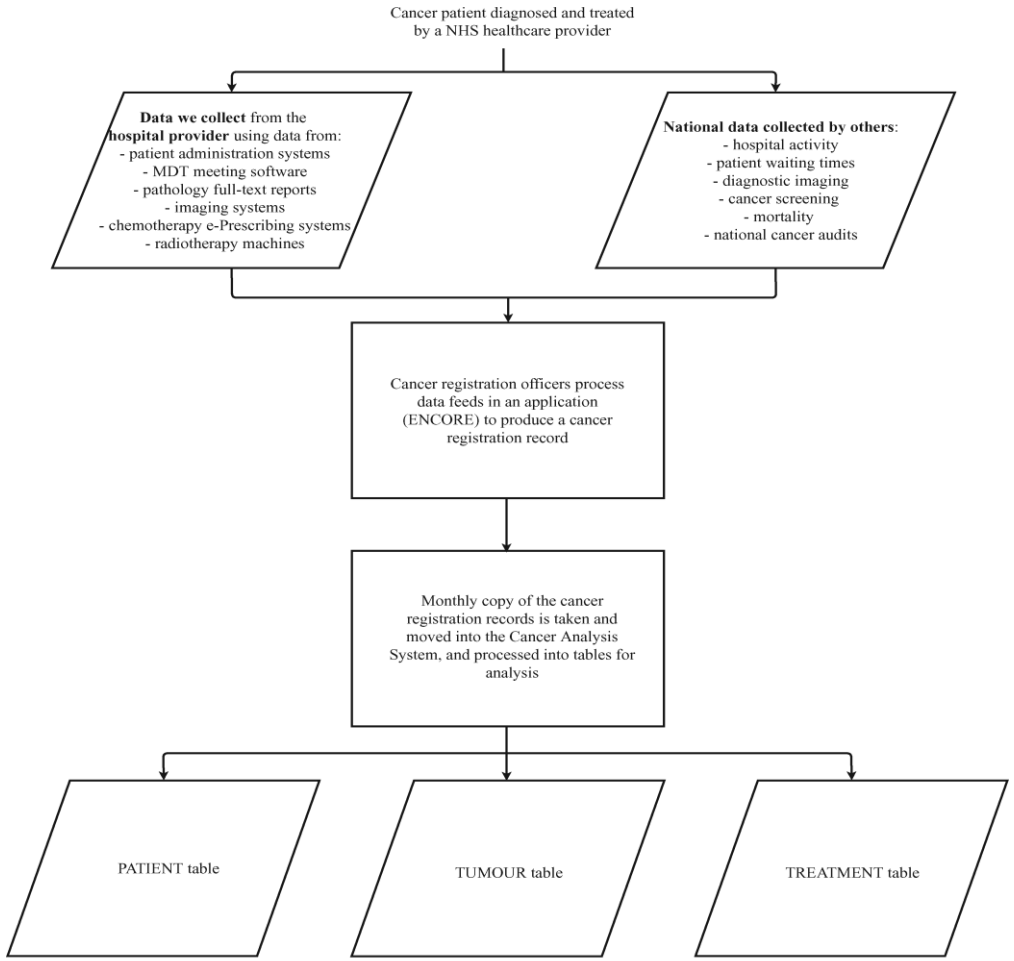
Specialised data registries: Cancer (1)

The National Cancer Registration and Analysis Service (NCRAS)

- Currently run by Public Health England (PHE) and migrating to NHS Digital
- Responsible for cancer registration in England to support cancer epidemiology, public health, service monitoring and research
- Collects information on over 300,000 cases of cancer each year
  - patient details (including name, address, age, sex, and date of birth)
  - type of cancer, how advanced at diagnosis, treatment received
- Multiple sources of data including
  - histopathology and haematology services
  - medical records
  - radiotherapy departments
  - hospices
  - independent hospitals
  - screening services
  - death certificates
  - GPs
  - other UK cancer registries

- Access requests via PHE Office for Data Release https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data

Figure 2. Summary of the data flows from the point a patient was diagnosed with cancer to the analytical views of data



ENCORE: English National Cancer Online Registration Environment
MDT: Multidisciplinary Team

Reproduced from Int J Epidemiol, Volume 49, Issue 1, February 2020, Pages 16–16h, https://doi.org/10.1093/ije/dyz076

Specialised data registries: Cancer (2)

Data Quality

- Data submitted are reviewed by skilled cancer registration officers (CROs)
- Some automated tools for data linkage and de-duplication of identical data sources
- Review includes manual extraction of data from text-based pathology reports
- CROs have detailed knowledge of cancer biology, coding and terminology
- CROs will, if necessary, seek additional source information from primary or secondary care via correspondence or direct interrogation of hospital radiology or electronic health records via remote secure access

Strengths

National coverage
Standardised coding (ICD-O3)
Timeliness (annual)
Minimal opt-out by patients

Limitations

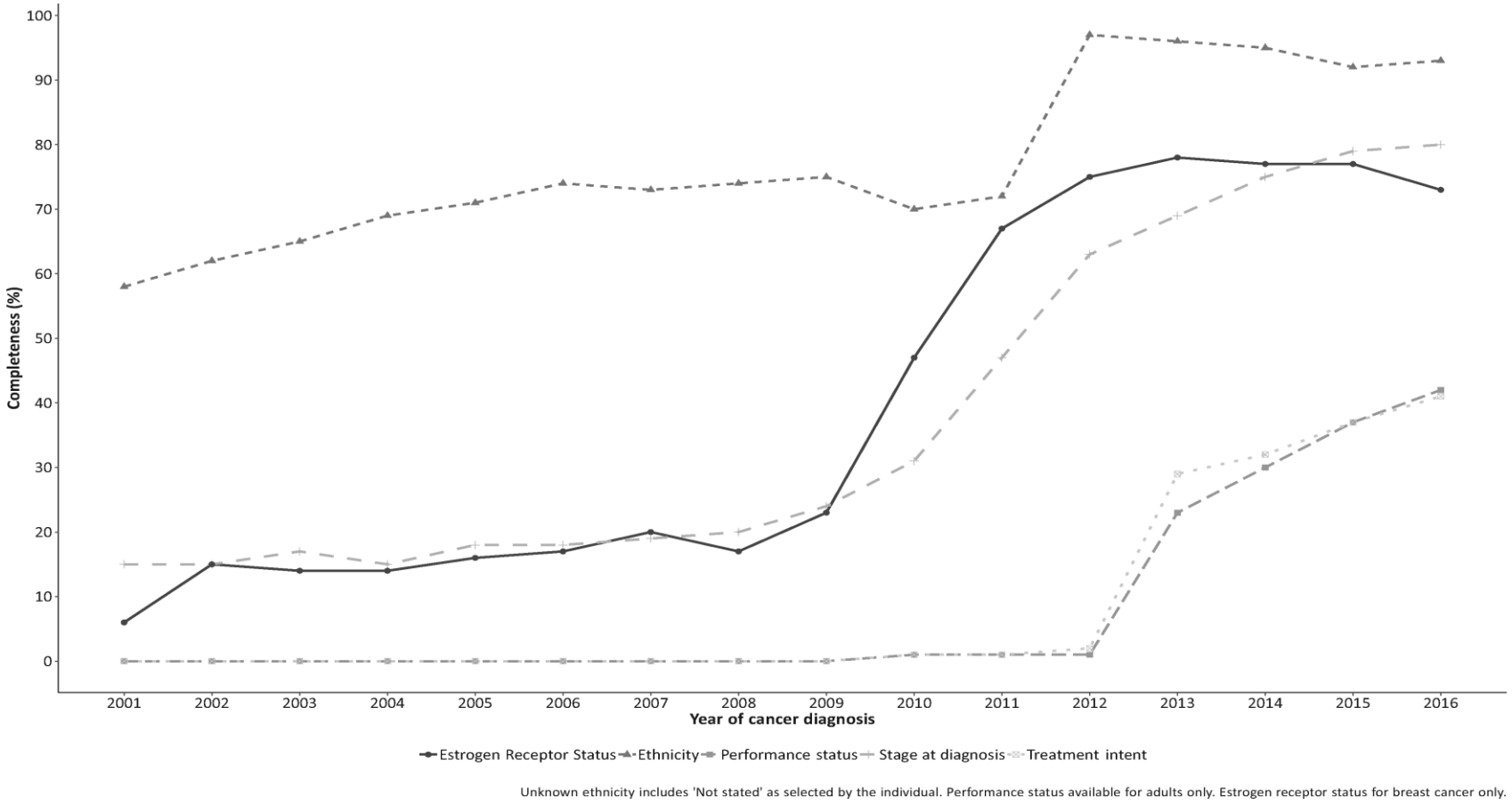
No national linkage to primary care data
Patient outcomes not well captured
Recurrence and progression poorly recorded
Limited information on lifestyle factors eg smoking, alcohol

References

- Data Resource Profile: National Cancer Registration Dataset in England International Journal of Epidemiology, 2020, 16–16h doi: 10.1093/ije/dyz076
- The National Cancer Registration and Analysis Service A guide to cancer data and working with us PHE March 2020

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/894416/National_Cancer_Registration_Guide-June-2020.pdf

Figure 3. Trend in data completeness (%) from 2001 to 2016 of the recording of estrogen receptor status for breast breast cancer patients, self-reported ethnicity, the performance status at diagnosis for adults, the registry-derived stage at cancer diagnosis, and the intent of treatment



Reproduced from Int J Epidemiol, Volume 49, Issue 1, February 2020, Pages 16–16h, https://doi.org/10.1093/ije/dyz076

Other specialised data registries in England

England

See
https://digital.nhs.uk/data-and-information/clinical-audits-and-registries

Examples include

- National Diabetes Audit Programme: Core, Inpatient, Foot care, Pregnancy in Diabetes, Harms

https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets

- Emergency Care Data Set
- Maternity and Children's Data Set

https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets

- National Congenital Anomaly and Rare Disease Registration Service (NCARDRS)

https://www.gov.uk/guidance/the-national-congenital-anomaly-and-rare-disease-registration-service-ncardrs

- National Cardiac Audit (NICOR) and MI/heart attack audit

https://www.nicor.org.uk/national-cardiac-audit-programme/myocardial-ischaemia-minap-heart-attack-audit/

Strengths and Limitations of using data collected about patients using hospital services for research

Pros

- National coverage
- Prospective
- Collected once, used many times
- Quality assured centrally
- Real world: records care in multiple settings not just specialist hospitals
- Can be linked to other national datasets

Limitations

- Data quality variable
- Collected for other purposes (billing) not research
- Coding is not undertaken by clinicians
- Not real-time
- No direct feedback to clinical care teams
- Lacks individual level data on events preceding and following hospital service use
- Lacks individual level data on wider determinants of health and life-style factors
- Limited data on outcomes and experiences of patients

Questions???

Key points

- Hospital Episode Statistics data are a valuable resource and widely used in research
- They provide national coverage of secondary care use and can be linked to other datasets
- Patients can have multiple attendances and admissions which need to be linked across provider, geography and time
- Not all diagnoses may be recorded as the primary purpose is for billing
- Coding may be variable in quality and completeness
- Specialised patient registries and audits provide additional depth of information which can enrich hospital episode data
- Patient experiences and outcomes of hospital care are not well recorded
- Data may be difficult to access and is not real-time

References

- Boyd A, Cornish R, Johnson L, Simmonds S, Syddall H, Westbury L, Cooper C, Macleod J. (2018). Understanding Hospital Episode Statistics (HES). Bristol, UK: University of Bristol.
- Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC) International Journal of Epidemiology, 2017, 1093–1093i doi: 10.1093/ije/dyx015
- Woodfield R, Grant I, UK Biobank Stroke Outcomes Group, UK Biobank Follow-Up and Outcomes Working Group, Sudlow CLM (2015) Accuracy of Electronic Health Record Data for Identifying Stroke Cases in Large-Scale Epidemiological Studies: A Systematic Review from the UK Biobank Stroke Outcomes Group. PLoS ONE 10(10): e0140533. doi:10.1371/journal.pone.0140533
- Rannikmäe et al Accuracy of identifying incident stroke cases from linked health care data in UK Biobank Neurology® 2020;95:e697-e707. doi:10.1212/WNL.0000000000009924
- The HES Processing Cycle and Data Quality. 2016, Health and Social Care Information Centre.
- Replacement of the HES Patient ID (HESID) 2009, The Health and Social Care Information Centre.
- Bishop C, Small N, Mason D, et al. Improving case ascertainment of congenital anomalies: findings from a prospective birth cohort with detailed primary care record linkage. BMJ Paediatrics Open 2017;1:e000171. doi:10.1136/bmjpo-2017-000171
- Doidge JC, Morris JK, Harron KL, Stevens S, Gilbert R. Prevalence of Down's Syndrome in England, 1998-2013: Comparison of linked surveillance data and electronic health records. Int J Popul Data Sci. 2020 Mar 19;5(1):1157. doi: 10.23889/ijpds.v5i1.1157

Tutorial

We will work through last assignment using RECORD checklist

You will be assigned an exercise to bring for discussion at next week's tutorial based on today's lecture

Understand the importance of using multiple data sources for research and for ascertainment of conditions and outcomes

This slide set will be made available to you on QMPlus together with a reading list